

Glossaire

Édition 202611

Copyright © EXIN Holding B.V. 2026. All rights reserved.
EXIN® is a registered trademark of EXIN.

No part of this publication may be reproduced, stored, utilized or transmitted in any form or by any means, electronic, mechanical, or otherwise, without the prior written permission from EXIN.

Introduction

Voici le glossaire de la certification EXIN AI Security Professional based on the OWASP AI Exchange, basée sur l'OWASP AI Exchange.

Il fournit les définitions des concepts de base répertoriés dans le guide de préparation et utilisés lors de l'examen, afin d'aider les candidats dans leur préparation.

Les définitions fournies sont conformes à celles de l'OWASP AI Exchange.

Glossaire

Les termes ci-dessous sont présentés dans le même ordre que dans la version anglaise, accompagnés de leurs définitions.

Terme	Définition
responsabilité *accountability*	Principe selon lequel les individus ou les organisations doivent rendre compte de leurs actions et de leurs décisions, et peuvent être tenus d'expliquer, de justifier ou d'assumer la responsabilité des conséquences.
attaque adversariale *adversarial attack*	Terme polysémique, parfois utilisé pour désigner toute attaque, et parfois pour désigner une tentative visant à créer ou à manipuler délibérément des données d'entrée, des données ou un modèle afin d'amener un système d'apprentissage automatique à produire des résultats incorrects, nuisibles ou choisis par l'attaquant.
test adversarial *adversarial testing*	Tests consistant à mener des attaques adversariales afin de mettre au jour des failles exploitables avant que de véritables attaquants ne le fassent.
IA agentique *agentic AI*	Systèmes construits autour d'agents IA capables de planifier, d'agir et d'interagir avec des outils, des données ou d'autres agents pour atteindre des objectifs avec une intervention humaine limitée.
IA (intelligence artificielle) *AI (artificial intelligence)*	Capacité d'un système informatique à effectuer des tâches qui requièrent généralement une intelligence de type humain, telles que le raisonnement, l'apprentissage, la reconnaissance de formes ou la création de contenu, que ce soit par des méthodes basées sur des règles ou par l'apprentissage automatique.
agent IA *AI agent*	Composant logiciel capable de décider et d'exécuter de manière autonome des actions, telles que l'appel de fonctions ou d'outils, afin d'atteindre un objectif, plutôt que de se contenter de produire une réponse.
actifs liés à l'IA *AI assets*	Éléments spécifiques à l'IA appartenant à un système ou à une organisation particulière et nécessitant une protection, tels que les données d'entraînement et les paramètres de modèles.
contournement des restrictions de l'IA (jailbreak) *AI jailbreak*	Attaque par injection de prompt contournant les restrictions de sécurité ou de contenu intégrées à un modèle afin de le forcer à produire un résultat qu'il était censé refuser.
système de management de l'IA (SMIA) *AI management system (AIMS)*	Cadre organisationnel composé de politiques et de processus, tel que défini par la norme ISO/IEC 42001, visant à régir l'IA de manière responsable et sécurisée.

Terme	Définition
modèle d'IA *AI model*	Artéfact mathématique ou informatique qui met en correspondance des entrées et des sorties, appris à partir de données ou défini par des règles.
fournisseur de modèle d'IA *AI model provider*	Une organisation ou une entité qui développe, entraîne ou fournit un modèle d'IA destiné à être utilisé ou à servir de base à d'autres.
principes de confidentialité de l'IA *AI privacy principles*	Ensemble de principes régissant le traitement des données à caractère personnel, couvrant des aspects tels que l'exactitude, le consentement, la minimisation, l'explicabilité, l'équité, la transparence et la limitation.
• **exactitude** • *accuracy*	Dans le contexte des données à caractère personnel : principe selon lequel les données à caractère personnel doivent être exactes et, si nécessaire, tenues à jour. Dans le contexte des modèles d'IA : degré de justesse des résultats fournis par un système d'IA.
• **consentement** • *consent*	Accord libre, spécifique, éclairé et univoque donné par la personne concernée pour le traitement de ses données à caractère personnel.
• **minimisation des données et limitation de la conservation** • *data minimization and storage limitation*	Principe consistant à ne collecter que les données nécessaires à une finalité spécifique et à ne pas les conserver plus longtemps que nécessaire pour cette finalité.
• **droits de la personne concernée** • *data subject rights*	Droits individuels relatifs aux données à caractère personnel, tels que l'accès, la rectification et la suppression.
• **explicabilité** • *explainability*	Principe selon lequel les personnes doivent être informées de la manière dont leurs données sont utilisées et, lorsque des décisions automatisées les concernent, être en mesure de comprendre les étapes qui ont conduit à ces décisions.
• **équité et licéité** • *fairness and lawfulness*	Principe selon lequel les données à caractère personnel doivent être traitées de manière légalement justifiée et sans porter injustement préjudice à des personnes ou à des groupes.
• **transparence** • *transparency*	Principe selon lequel les personnes doivent être informées de la manière dont leurs données sont collectées, utilisées, partagées et conservées et, lorsque des décisions automatisées les concernent, être en mesure de comprendre la méthode suivie pour parvenir à ces décisions.
• **limitation de l'utilisation et spécification de la finalité** • *use limitation and purpose specification*	Le principe de protection de la vie privée selon lequel les données à caractère personnel collectées dans un but donné ne doivent pas être simplement réutilisées pour entraîner un modèle d'IA à des fins différentes et incompatibles sans fondement juridique approprié.

Terme	Définition
red teaming de l'IA (exercice de test adversarial) *AI red teaming*	Exercice de test structuré et adversarial dans lequel les testeurs se comportent comme de véritables attaquants afin de mettre au jour les failles d'un système en matière de sécurité, de sûreté ou d'éthique.
sécurité de l'IA *AI security*	Discipline visant à protéger les systèmes d'IA, leurs données et leur infrastructure sous-jacente contre les menaces de sécurité qui exploitent les caractéristiques spécifiques à l'IA, en plus des risques de sécurité classiques liés à la cybersécurité.
gouvernance de la sécurité de l'IA *AI security governance*	Ensemble des structures organisationnelles, des politiques et des processus de contrôle qui garantissent que les risques de sécurité liés à l'IA sont identifiés, gérés et font l'objet d'une responsabilité clairement attribuée.
approche des tests de sécurité de l'IA *AI security testing approach*	Une méthodologie structurée pour planifier et exécuter des tests de sécurité sur un système d'IA, couvrant le périmètre, les techniques et les critères de réussite.
système d'IA *AI system*	Système complet articulé autour d'un ou plusieurs modèles d'IA, comprenant ses pipelines de données, ses interfaces et son infrastructure de soutien.
fournisseur de système d'IA *AI system provider*	Organisation qui développe ou assemble un système d'IA et le met à disposition pour son déploiement ou son utilisation.
anonymisation *anonymization*	Traitement irréversible des données à caractère personnel de manière à ce que la personne concernée ne puisse plus être identifiée, même indirectement.
API (interface de programmation d'application) *API (application programming interface)*	Ensemble défini de protocoles et d'outils permettant à différents logiciels ou composants de communiquer et d'échanger des fonctionnalités.
audit *audit*	Examen indépendant d'un système, d'un processus ou d'une organisation visant à vérifier la conformité aux normes, politiques ou réglementations.
données d'augmentation *augmentation data*	Données externes ajoutées au contexte d'un modèle à un moment donné de son cycle de vie afin d'améliorer ou d'étendre ses résultats, telles que des documents récupérés pour la génération augmentée par récupération (RAG).
contrôle automatisé *automated oversight*	Surveillance ou contrôle d'un système effectué par des outils d'automatisation plutôt que par des personnes.
automatisation *automation*	Utilisation de la technologie pour effectuer des tâches ou des processus avec une intervention humaine réduite.
biais *bias*	Un biais systématique dans les données, les décisions ou les résultats, souvent introduit involontairement par des choix relatifs aux données ou à la conception. Selon les cas, ce biais peut conduire à un traitement injuste ou inexact de certains groupes ou cas.

Terme	Définition
rayon d'impact *blast radius*	L'étendue des dommages, des compromissions ou des divulgations qui résulteraient d'une attaque réussie ou d'une défaillance d'un actif, d'un composant ou d'un identifiant particulier.
agent conversationnel (chatbot) *chatbot*	Application logicielle qui interagit avec les utilisateurs par le biais d'une conversation textuelle ou vocale, en automatisant des réponses qui nécessiteraient autrement l'intervention d'un être humain.
directeur/directrice de l'IA (CAIO) *chief AI officer (CAIO)*	Fonction au sein d'une organisation chargée de la stratégie globale en matière d'IA, de la gouvernance et de la supervision des risques.
responsable de la sécurité des systèmes d'information (RSSI) *Chief Information Security Officer (CISO)*	Cadre supérieur chargé de la stratégie et du programme globaux d'une organisation en matière d'information et de cybersécurité.
code *code*	Ensemble d'instructions, rédigées dans un langage de programmation, qui mettent en œuvre la logique d'un logiciel.
conformité *compliance*	Respect des lois, réglementations, normes et politiques internes applicables régissant les activités d'une organisation.
confidentialité *confidentiality*	Propriété selon laquelle les informations ne sont accessibles qu'aux personnes autorisées à y accéder. L'un des trois objectifs fondamentaux de la sécurité de l'information, au même titre que l'intégrité et la disponibilité.
fuite de configuration *configuration leak*	Divulgation involontaire des détails de configuration d'un système, tels que les paramètres ou les identifiants, à des tiers non autorisés.
droit d'auteur *copyright*	Forme de protection juridique accordée aux créateurs d'œuvres originales, leur conférant des droits exclusifs sur la reproduction, la distribution et l'utilisation de ces œuvres.
fuite de données *data leak*	La divulgation non autorisée ou involontaire de données à des personnes qui ne devraient pas y avoir accès.
protection des données *data protection*	Dans le contexte des données à caractère personnel, ensemble des mesures et pratiques visant à protéger ces données contre tout accès, toute utilisation ou toute divulgation non autorisés.
analyse d'impact relative à la protection des données (DPIA) *data protection impact assessment (DPIA)*	Processus structuré, requis par la législation sur la protection des données, visant à identifier et à atténuer les risques liés à la protection de la vie privée avant de procéder à un traitement de données susceptible d'entraîner un risque élevé pour les personnes.
sécurité des données *data security*	Protection des données contre tout accès non autorisé, toute modification ou toute perte tout au long de leur cycle de vie.

Terme	Définition
personne concernée *data subject*	Personne physique vivante, identifiée ou identifiable, à laquelle se rapportent des données à caractère personnel.
déni de service (DoS) *denial-of-service (DoS)*	Attaque visant à rendre un système ou un service indisponible pour les utilisateurs légitimes, généralement en le saturant de trafic ou de requêtes.
déni de portefeuille (DoW) *denial-of-wallet (DoW)*	Attaque visant à faire grimper délibérément les coûts d'exploitation d'une cible, par exemple en déclenchant des requêtes excessives et gourmandes en ressources sur un service facturé à l'utilisation, causant ainsi un préjudice financier plutôt qu'une indisponibilité.
déployeur *deployer*	Personne ou organisation qui met un système en service, par opposition à celle qui l'a développé.
phase de développement *development-time*	Phase du cycle de vie d'un système durant laquelle celui-ci est conçu, construit, entraîné ou testé, par opposition à la phase de déploiement et d'exploitation.
menaces de sécurité en phase de développement *development-time security threats*	Risques de sécurité pouvant survenir lors de la conception, de la construction ou de l'entraînement d'un système, tels qu'un code compromis, des composants comportant des portes dérobées ou des données corrompues.
• **empoisonnement des données en phase de développement** • *data poisoning development-time*	Action consistant à corrompre ou à manipuler délibérément les données utilisées pour entraîner un modèle d'apprentissage automatique, afin que le modèle résultant acquière un comportement erroné ou contrôlé par un attaquant.
• **empoisonnement du modèle en phase de développement** • *model poisoning development-time*	Altération directe des paramètres, de l'architecture du modèle ou du processus d'entraînement afin d'y intégrer un comportement caché ou malveillant.
• **fuite de données sensibles en phase de développement** • *sensitive data leak development-time*	La divulgation involontaire de données sensibles, telles que des données sources, des identifiants ou des secrets de configuration, au cours du processus de développement d'un système.
plongement vectoriel (embedding) *embedding*	Représentation vectorielle numérique de données, telles que du texte ou des images, qui traduit leur signification sémantique afin que les modèles d'apprentissage automatique puissent les traiter, les comparer ou les rechercher.
chiffrement *encryption*	Processus consistant à coder des données de manière à ce qu'elles ne puissent être lues que par les parties disposant de la clé de déchiffrement appropriée, utilisé pour protéger la confidentialité des données.

Terme	Définition
principes éthiques *ethical principles*	Ensemble de valeurs, telles que l'équité, l'honnêteté et la responsabilité, qui définissent ce qui est considéré comme un comportement juste ou acceptable dans un domaine ou une activité particulière.
Union européenne (UE) *European Union (EU)*	Union politique et économique des États membres européens qui adopte des lois et des réglementations, telles que le RGPD et le règlement sur l'IA, contraignantes pour ses membres.
G.U.A.R.D (Gouverner, Comprendre, Adapter, Réduire, Démontrer) *G.U.A.R.D (Govern, Understand, Adapt, Reduce, Demonstrate)*	Cadre en cinq étapes proposé par l'OWASP AI Exchange (Gouverner, Comprendre, Adapter, Réduire, Démontrer) pour structurer l'approche d'une organisation en matière de gestion des risques de sécurité de l'IA.
Règlement général sur la protection des données (RGPD) *General Data Protection Regulation (GDPR)*	Règlement de l'Union européenne qui définit les règles relatives à la protection et au traitement des données à caractère personnel des individus.
IA générative *generative AI*	Systèmes d'IA qui créent de nouveaux contenus, tels que du texte, des images ou du code, généralement en apprenant des schémas à partir de vastes ensembles de données d'entraînement.
détournement *hijacking*	Attaque au cours de laquelle un attaquant prend le contrôle non autorisé d'un système, d'une session, d'une ressource ou d'un canal de communication et le détourne à ses propres fins.
contrôle humain *human oversight*	Implication de personnes dans la surveillance, la révision ou l'intervention sur les décisions et les actions d'un système automatisé ou autonome.
phase d'inférence *inference-time*	Étape du cycle de vie d'un modèle d'apprentissage automatique au cours de laquelle celui-ci est activement utilisé pour produire des résultats à partir de nouvelles données d'entrée, par opposition à la phase d'entraînement.
système de management de la sécurité de l'information (SMSI) *information security management system (ISMS)*	Cadre systématique de politiques, de processus et de contrôles, tel que défini dans la norme ISO/IEC 27001, destiné à gérer les risques de sécurité de l'information d'une organisation.
données d'entrée *input data*	Données fournies à un système ou à un processus afin de produire un résultat. Dans le contexte de l'IA, il s'agit des données utilisées comme entrée pour le modèle d'IA.

Terme	Définition
menaces de sécurité liées aux données d'entrée *input security threats*	Risques de sécurité découlant des données ou du contenu fournis à un système en tant qu'entrée, tels que des contenus malveillants ou manipulés destinés à altérer son comportement.
• **évasion** • *evasion*	Attaque consistant à créer des données d'entrée malveillantes, par opposition à des instructions, spécialement conçues pour contourner un système de détection ou de classification, afin de le pousser à produire un résultat erroné. Par exemple, un panneau de signalisation subtilement modifié qu'un modèle de vision interprète de manière erronée.
• **exfiltration du modèle** • *model exfiltration*	Extraction d'une copie ou d'un équivalent fonctionnel d'un modèle propriétaire, souvent par le biais de requêtes répétées, sans autorisation.
• **injection de prompt** • *prompt injection*	Attaque consistant à manipuler les instructions données à un modèle d'IA générative, par le biais du prompt ou du contenu injecté, afin de contourner le comportement prévu.
• **épuisement des ressources** • *resource exhaustion*	Attaque ou situation consistant à consommer des ressources de calcul, de mémoire ou de stockage au point de dégrader ou d'empêcher l'accès au service pour les utilisateurs légitimes, ou d'entraîner un impact financier important.
• **divulgation de données sensibles par l'utilisation** • *sensitive data disclosure through use*	Divulgation de données sensibles d'entraînement ou d'autres données par le biais d'une utilisation normale ou abusive du système.
vérification de l'intégrité *integrity verification*	Vérification qu'une donnée, un code ou tout autre élément numérique n'a pas été altéré ou falsifié, généralement à l'aide de techniques telles que les sommes de contrôle, les hachages ou les signatures numériques.
propriété intellectuelle (PI) *intellectual property (IP)*	Créations de l'esprit protégées par la loi, telles que les secrets d'affaires, les brevets, les œuvres protégées par le droit d'auteur et les marques, sur lesquelles leur créateur ou leur propriétaire détient des droits exclusifs.
ISO/IEC 42001 *ISO/IEC 42001*	Norme internationale spécifiant les exigences relatives à un système de management de l'IA visant à régir l'IA de manière responsable et sécurisée.
grand modèle de langage (LLM) *large language model (LLM)*	Modèle d'apprentissage automatique entraîné sur de vastes quantités de texte, capable d'interpréter et de générer du langage naturel.
fuite *leakage*	Divulgation non autorisée ou involontaire d'informations qui auraient dû rester confidentielles ou protégées.

Terme	Définition
apprentissage automatique *machine learning*	Branche de l'IA dans laquelle les systèmes d'IA apprennent des schémas à partir de données plutôt que d'être explicitement programmés avec des règles.
inférence d'appartenance *membership inference*	Attaque visant à déterminer si un enregistrement spécifique faisait partie des données d'entraînement d'un modèle, ce qui peut révéler des informations sensibles sur des individus.
architecture du modèle *model architecture*	Conception structurelle d'un modèle, comprenant ses composants et la manière dont ils sont reliés entre eux.
inversion de modèle *model inversion*	Attaque consistant à reconstituer ou à estimer des données d'entraînement sensibles en analysant les résultats d'un modèle.
fuite de modèle *model leak*	Divulgation non autorisée ou vol des paramètres, des poids ou de l'architecture du modèle.
open source *open source*	Logiciel ou autre artéfact technique dont le code source ou les composants sous-jacents sont mis à la disposition du public afin que chacun puisse les utiliser, les inspecter, les modifier et les redistribuer.
données de sortie *output data*	Résultat produit par un système ou un processus en réponse à une entrée donnée. Dans le contexte de l'IA, il s'agit généralement de la sortie d'un modèle d'IA.
assainissement des sorties *output sanitization*	Filtrage ou nettoyage de la sortie d'un système avant qu'elle ne soit transmise ou affichée, afin de supprimer tout contenu nuisible, sensible ou malveillant.
validation des sorties *output validation*	Vérification de la sortie d'un système par rapport à des critères définis afin de s'assurer qu'elle est correcte, sûre et adaptée à l'usage prévu avant qu'elle ne soit exploitée ou transmise en aval.
données à caractère personnel *personal data*	Toute information se rapportant à une personne physique vivante, identifiée ou identifiable.
IA prédictive *predictive AI*	Systèmes d'IA conçus pour prévoir des résultats ou classer des données en fonction de schémas appris à partir de données historiques, par opposition à la génération de nouveaux contenus.
protection de la vie privée *privacy*	Droit d'une personne à contrôler la manière dont ses informations personnelles sont collectées, utilisées et partagées.
protection de la vie privée par défaut *privacy by default*	Le principe selon lequel les paramètres par défaut d'un système doivent constituer l'option la plus protectrice de la vie privée parmi celles disponibles, sans nécessiter d'action supplémentaire de la part de l'utilisateur, un concept complémentaire à celui de la « protection de la vie privée par conception ».

Terme	Définition
protection de la vie privée par conception *privacy by design*	Intégration des mesures de protection de la vie privée dans l'architecture et les processus d'un système dès les premières étapes de son développement, plutôt que de leur ajout a posteriori.
analyses d'impact relatives à la vie privée (PIA) *privacy impact assessments (PIA)*	Évaluation visant à identifier et à traiter les risques en matière de protection de la vie privée qu'un projet, un système ou un processus peut présenter pour les personnes.
incident lié à la vie privée *privacy incident*	Événement au cours duquel des données à caractère personnel sont traitées, consultées ou divulguées d'une manière qui enfreint les mesures de protection de la vie privée.
traitement (des données à caractère personnel) *processing (of personal data)*	Toute opération effectuée sur des données à caractère personnel, telle que la collecte, le stockage, l'utilisation ou la suppression.
prompt *prompt*	Texte ou instructions fournis à un modèle d'IA générative afin d'obtenir une réponse.
pseudonymisation *pseudonymization*	Remplacement des informations d'identification contenues dans les données par des identifiants artificiels, de sorte que des informations supplémentaires, conservées séparément, soient nécessaires pour réidentifier la personne.
accessible au public *publicly accessible*	Données, systèmes ou ressources accessibles à toute personne sans authentification ni autorisation particulière.
modèle prêt à l'emploi *ready-made model*	Modèle pré-entraîné obtenu auprès d'un tiers et utilisé tel quel ou après adaptation, plutôt que d'être entraîné à partir de zéro.
risque résiduel *residual risk*	Risque qui subsiste après la mise en œuvre de contrôles de sécurité et de mesures d'atténuation.
IA responsable *responsible AI*	Une approche de développement et d'exploitation des systèmes d'IA qui tient compte de l'éthique, de l'équité, de la sûreté et de la sécurité au même titre que des performances.
génération augmentée par récupération (RAG) *retrieval-augmented generation (RAG)*	Technique dans laquelle un modèle d'IA générative récupère des documents ou des données externes pertinents lors de la phase d'inférence afin d'étayer et d'enrichir ses réponses.
risque *risk*	Possibilité qu'une menace exploite une vulnérabilité et cause un préjudice, généralement évaluée en termes de probabilité et d'impact.
analyse des risques *risk analysis*	Processus consistant à identifier et à évaluer les risques afin de déterminer leur probabilité et leur impact potentiel.

Terme	Définition
niveaux de risque des systèmes d'IA selon le règlement sur l'IA *risk levels for AI systems based on AI Act*	Classification des systèmes d'IA en différents niveaux de risque réglementaire prévue par le règlement européen sur l'IA, qui détermine les obligations légales applicables à un système donné.
• **risque minimal ou nul** • *minimal or no risk*	Niveau défini par le règlement sur l'IA pour les systèmes d'IA considérés comme présentant un risque négligeable pour la santé, la sécurité ou les droits fondamentaux, et soumis à une réglementation faible, voire inexistante.
• **risque limité** • *limited risk*	Niveau défini par le règlement sur l'IA pour les systèmes d'IA présentant un certain risque, généralement soumis à des obligations spécifiques en matière de transparence, telles que la mention indiquant que le contenu est généré par l'IA.
• **haut risque** • *high risk*	Niveau défini par le règlement sur l'IA pour les systèmes d'IA susceptibles d'avoir un impact significatif sur la santé, la sécurité ou les droits fondamentaux, soumis à des obligations strictes avant et après leur déploiement.
• **risque inacceptable** • *unacceptable risk*	Niveau défini par le règlement sur l'IA pour les pratiques d'IA considérées comme présentant un risque si grave qu'elles sont purement et simplement interdites.
gestion des risques *risk management*	Processus continu consistant à identifier, évaluer, traiter et surveiller les risques afin de les maintenir à des niveaux acceptables.
atténuation des risques *risk mitigation*	Sélection et mise en œuvre de mesures visant à réduire la probabilité ou l'impact d'un risque identifié.
cotation du risque *risk rating*	Note ou classification attribuée à un risque en fonction de sa probabilité et de son impact évalués.
phase d'exécution *runtime*	Phase du cycle de vie d'un système durant laquelle celui-ci est déployé et fonctionne activement, en traitant des données réelles dans un environnement de production.

Terme	Définition
menaces de sécurité en phase d'exécution *runtime security threats*	Risques de sécurité qui surviennent lorsqu'un système fonctionne dans un environnement de production réel, par opposition à la phase de développement.
• **manipulation des données d'augmentation** • *augmentation data manipulation*	Altération des données externes qu'un système récupère ou injecte lors de la phase d'exécution, telles que les sources de génération augmentée par récupération (RAG), afin d'influencer sa sortie.
• **fuite directe des données d'augmentation** • *direct augmentation data leak*	Divulgation des données d'augmentation par un accès direct non autorisé, que ce soit en transit ou au repos.
• **fuite directe du modèle en phase d'exécution** • *direct runtime model leak*	Divulgation non autorisée des paramètres ou de l'architecture du modèle par le biais d'un accès direct non autorisé, en transit ou au repos.
• **empoisonnement direct du modèle en phase d'exécution** • *direct runtime model poisoning*	Manipulation par un attaquant du comportement d'un modèle déployé via un accès direct non autorisé lors de la phase d'exécution, en transit ou au repos.
• **fuite des données d'entrée** • *input data leak*	Divulgation non autorisée de données fournies en entrée à un système.
• **sortie contenant une injection conventionnelle** • *output containing conventional injection*	Sortie d'un système contenant elle-même du contenu malveillant, tel que des fragments de script ou de requête, qui déclenche une attaque par injection conventionnelle, telle qu'une injection SQL ou un cross-site scripting (XSS), lorsqu'elle est transmise à un système en aval.
sûreté d'un système d'IA *safety of an AI system*	Propriété selon laquelle un système d'IA ne cause pas de préjudice inacceptable aux personnes, aux biens ou à l'environnement du fait de son fonctionnement.
attaque (de sécurité) *(security) attack*	Action délibérée menée par un attaquant dans le but de compromettre la confidentialité, l'intégrité ou la disponibilité d'un système ou de ses actifs.
violation de sécurité *(security) breach*	Incident au cours duquel les contrôles de sécurité sont contournés ou échouent, entraînant un accès non autorisé, une divulgation ou des dommages.

Terme	Définition
contrôles de sécurité *security controls*	Mesures techniques, procédurales ou organisationnelles mises en place pour réduire les risques de sécurité.
• **contrôles de limitation des données sensibles** • *controls for sensitive data limitation*	Contrôles de sécurité visant spécifiquement à restreindre la collecte, la divulgation ou la conservation des données sensibles.
• **contrôles visant à limiter les effets des comportements indésirables** • *controls to limit the effect of unwanted behavior*	Contrôles de sécurité conçus pour contenir ou réduire l'impact d'un comportement involontaire ou nuisible d'un système une fois qu'il se produit.
• **contrôles de gouvernance** • *governance controls*	Contrôles au niveau organisationnel, tels que les politiques, les structures de supervision et les mécanismes de responsabilité, qui gèrent les risques à un niveau stratégique.
incident de sécurité *security incident*	Événement indiquant une éventuelle violation de la politique de sécurité, une défaillance des contrôles de sécurité ou la compromission d'un actif.
risques de sécurité *security risks*	Possibilité que des menaces de sécurité exploitent des vulnérabilités et causent un préjudice.
tests de sécurité *security testing*	Processus consistant à évaluer un système afin d'identifier ses failles de sécurité avant qu'elles ne puissent être exploitées.
menaces de sécurité *security threats*	Ensemble complet des moyens par lesquels la confidentialité, l'intégrité ou la disponibilité d'un système peuvent être compromises. Dans le contexte des systèmes d'IA, cela inclut les menaces liées aux données d'entrée, les menaces liées à la phase de développement et les menaces conventionnelles liées à la phase d'exécution.
données sensibles *sensitive data*	Données nécessitant une protection particulière en raison du préjudice potentiel qu'elles pourraient causer si elles étaient divulguées, telles que les données à caractère personnel, les informations confidentielles ou les informations classifiées.
fournisseur de service *service provider*	Un tiers qui fournit un service, tel que l'hébergement, l'infrastructure ou un logiciel, dont dépend une organisation ou un système.
code source *source code*	Instructions lisibles par l'homme, écrites dans un langage de programmation, qui constituent un logiciel.
fuite du code source *source code leak*	Divulgation non autorisée du code source d'un système à des tiers non autorisés.

Terme	Définition
attaque éponge *sponge attack*	Attaque consistant à alimenter un modèle avec des données d'entrée conçues pour maximiser son temps de calcul ou sa consommation de ressources, afin d'en dégrader les performances ou la disponibilité.
partie prenante *stakeholder*	Toute personne ou organisation ayant un intérêt dans un projet, un système ou une décision, ou qui est concernée par ceux-ci.
jeune pousse *startup*	Entreprise en phase de démarrage, généralement fondée pour développer et déployer à grande échelle un nouveau produit, une nouvelle technologie ou un nouveau modèle économique.
langage de requête structuré (SQL) *structured query language (SQL)*	Langage standard utilisé pour créer, interroger et gérer les données stockées dans des bases de données relationnelles.
altération *tampering*	Modification non autorisée de données, de code ou d'un système visant à modifier son comportement ou à compromettre son intégrité.
tiers *third party*	Toute organisation ou personne externe, autre que le fournisseur et l'utilisateur principaux, qui est néanmoins impliquée dans un système ou qui a une incidence sur celui-ci, par exemple via sa chaîne d'approvisionnement.
modélisation des menaces *threat modeling*	Processus structuré visant à identifier les menaces potentielles, les vulnérabilités et les voies d'attaque d'un système, utilisé pour hiérarchiser les mesures de sécurité.
jeton *token*	Unité de base de texte, telle qu'un mot, une partie de mot ou un caractère, en laquelle le langage est décomposé pour être traité par un modèle de langage. Dans le contexte des services d'IA dans le cloud, les jetons constituent également l'unité de facturation.
tokenisation *tokenization*	Processus consistant à diviser un texte en unités plus petites en vue de leur traitement par un modèle de langage, ou à remplacer des éléments de données sensibles par des valeurs de remplacement non sensibles afin de protéger les données d'origine.
outil *tool*	Dans le contexte de l'IA agentique, une fonction externe, une API (interface de programmation d'application) ou une capacité qu'un agent IA peut invoquer pour effectuer des actions allant au-delà de la génération de texte.
données d'entraînement *training data*	Données utilisées pour entraîner un modèle d'apprentissage automatique à reconnaître des schémas ou à accomplir sa tâche.
IA digne de confiance *trustworthy AI*	IA conçue et exploitée de manière à être conforme à la loi, éthique et techniquement robuste, notamment en termes de sécurité et de sûreté.

Terme	Définition
comportement indésirable *unwanted behavior*	Tout résultat, décision ou action d'un système qui s'écarte de son comportement prévu, sûr ou attendu.
vulnérabilité *vulnerability*	Faiblesse d'un système, d'un processus ou d'un contrôle susceptible d'être exploitée par une menace pour causer un préjudice.

Certified for what's next

Contacter EXIN

www.exin.com